DNA sequence of the promoter region of the *ompC* gene and the amino acid sequence of the signal peptide of pro-OmpC protein of *Escherichia coli*

Takeshi Mizuno, Mei-Yin Chou and Masayori Inouye

Department of Biochemistry, State University of New York at Stony Brook, Stony Brook, NY 11794, USA

Received 12 November 1982

The osmoregulated *ompC* gene of *Escherichia coli* was cloned and the DNA sequence of a fragment encompassing the promoter region and a portion of the coding region was determined. There were no obvious homologies in the DNA sequences of the promoter regions of the *ompC* and *ompF* genes, in contrast to those of the coding regions of the two genes, both of which code for the matrix porins (major outer membrane proteins) and form passive diffusion pores. The amino acid sequence of the signal peptide of pro-OmpC protein was also deduced from the DNA sequence

Pro-OmpC protein *Signal peptide sequence* (Escherichia coli) ompC *gene* *Matrix porin* *Promoter region sequence*

1. INTRODUCTION

The matrix proteins (major outer membrane proteins) of *Escherichia coli,* are known to consist of OmpF and OmpC proteins. Both proteins form passive diffusion pores across the outer membrane, which are essential for the up-take of nutrients such as sugars and amino acids (review [1]).

These two proteins are coded for by two independent structural genes, the *ompF* gene and the *ompC* gene, which map at 21 and 47 min, respectively, on the *E. coli* chromosome (review [2]). The expression of these genes is controlled by the osmolarity of the culture medium; the *ompF* gene is preferentially expressed in low osmolarity, while the *ompC* gene is expressed under conditions of high osmolarity [3]. This osmoregulation of the *ompF* and *ompC* genes has been shown to be controlled by another unlinked gene, called the *ompB* operon, which maps at 74 min on the *E. coli* chromosome [2]. The DNA sequence of the entire *ompB* operon has been recently determined and its gene products have been identified [4,5].

To fully understand the molecular mechanism of the osmoregulation of the *ompF* and *ompC* genes, it is essential to isolate these genes, and to determine the structures of their promoters, with which the *ompB* gene products are considered to interact. The DNA sequence of the *ompF* promoter region has been determined in [6].

Here, we report the cloning of the entire *ompC* gene and the DNA sequence of it promoter region. The amino acid sequence of the signal peptide of the pro-OmpC protein has been deduced from the DNA sequence. These features will be compared with those of the *ompF* gene.

2. MATERIALS AND METHODS

Restriction endonucleases were obtained from Bethesda Res. Lab. and New England BioLabs. T4 DNA ligase was from Bethesda Res. Lab. DNA polymerase I Klenow fragment was from New England Nuclear. [α-^{32}P]dNTPs (2000–3000 Ci/mmol) were purchased from Amersham International.

2.1. *Bacterial strains and bacteriophages*

The following *E. coli* strains and phages were used: *E. coli* T19 *recA* (F^- *tsx*-354 *ompB*) [7],

Published by Elsevier Biomedical Press
00145793/83/0000–0000/$3.00

E. coli W620 *recA* ($F^-\lambda^-$ *thi*-1 *pyr*D36 *glt*A6 *gal*K30 *str*AD9 *supE*44 *rel*A1?), *E. coli* W620-78 *ompC*::Mu which was derived from W620 by integrating phage Mu into the *ompC* gene [7], phage λp10-25 (*ompC–lacZ*$^+$ 10–25 *lacY*$^+$ *lacA*$^+$; a gift from T.J. Silhavy [8]), and phage TuIb (from J. Foulds).

2.2. *Preparation of DNA and restriction endonuclease digestion*

Plasmid DNA, chromosomal DNA, and phage DNA were prepared as in [7]. Electrophoresis of DNA fragments [9] and restriction enzyme digestions [10] were carried out as described.

2.3. *DNA sequence determination*

Restriction fragments were labeled at their 3′-ends using [α-^{32}P]dNTPs and DNA polymerase I Klenow fragment [11]. Singly end-labeled DNA fragments were obtained either by digestion with a second restriction enzyme or electrophoretic strand separation [12]. DNA sequences were determined as in [12,13].

2.4. *Other methods*

Conditions for ligation and transformation of plasmids were as in [10]. DNA used for Southern blot hybridization [14] was nick translated as in [15].

3. RESULTS

Originally we attempted to determine the DNA sequence of the *ompC* promoter region and the region corresponding to the 5′-end of the *ompC* mRNA by directly sequencing a DNA fragment derived from λp10-25 DNA carrying an *ompC–lacZ* fusion [8]. However, this attempt was unsuccessful since it was found that this fusion phage contained a fragment of Mu DNA between the *ompC* gene and the *lacZ* gene as shown in fig.1A. The DNA sequence of the *Pvu*II–*Hpa*I fragment obtained

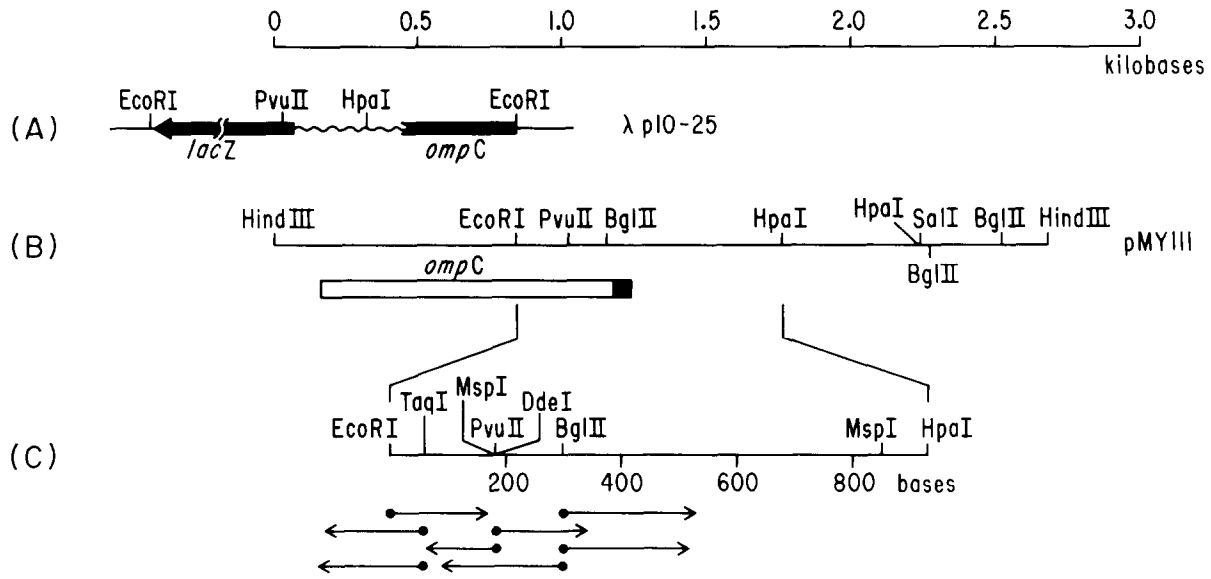

Fig.1. Cloning of the *ompC* gene and the sequencing strategy. (A) The 3.7-kb *Eco*RI fragment in λp10-25 carrying an *ompC–lacZ* gene fusion. The *lacZ* gene (←) was assigned in the 2.9-kb *Eco*RI–*Pvu*II region on the basis of DNA sequencing results. The region indicated by a waved line (~~~) was found to be derived from Mu phage DNA (see text). (B) The 2.7-kb *Hin*dIII fragment from chromosomal DNA, which was cloned into pBR322 (the plasmid was designated as pMY111). The *Eco*RI site corresponds to the *Eco*RI site of the *Hpa*I–*Eco*RI fragment in λp10-25. The approximate coding region for the *ompC* gene (▭) is shown under the restriction map. The closed box (▬) indicates the coding region of the signal peptide sequence. (C) Sequencing strategy: (→) extent of sequences determination using 3′-end (●) labeled fragments.

from λp10-25 (see fig.1A) indicated that this fragment carried a portion of the 5'-end of the *lacZ* gene and a sequence which did not resemble the known NH_2-terminal amino acid sequence of OmpC protein [16] (not shown). Therefore, we concluded that this region was derived from Mu phage DNA which was employed in the construction of the *ompC–lacZ* gene fusion [8].

Therefore, in order to isolate the *ompC* gene, we decided to use the upstream *Hpa*I–*Eco*RI fragment (~500 bp long) as a probe (see fig.1A), since this region was considered to contain a part of the *ompC* gene.

Chromosomal DNA from *E. coli* W620 was digested with several restriction enzymes, and hybridized with the ^{32}P-labeled *Hpa*I–*Eco*RI fragment. A 2.7-kb *Hin*dIII, a 4.7-kb *Hpa*I. and a 2.0-kb *Eco*RI fragment were found to hybridize with the probe (not shown). The 2.7-kb *Hin*dIII fragment was then cloned into pBR322. *E. coli* T19 *recA*

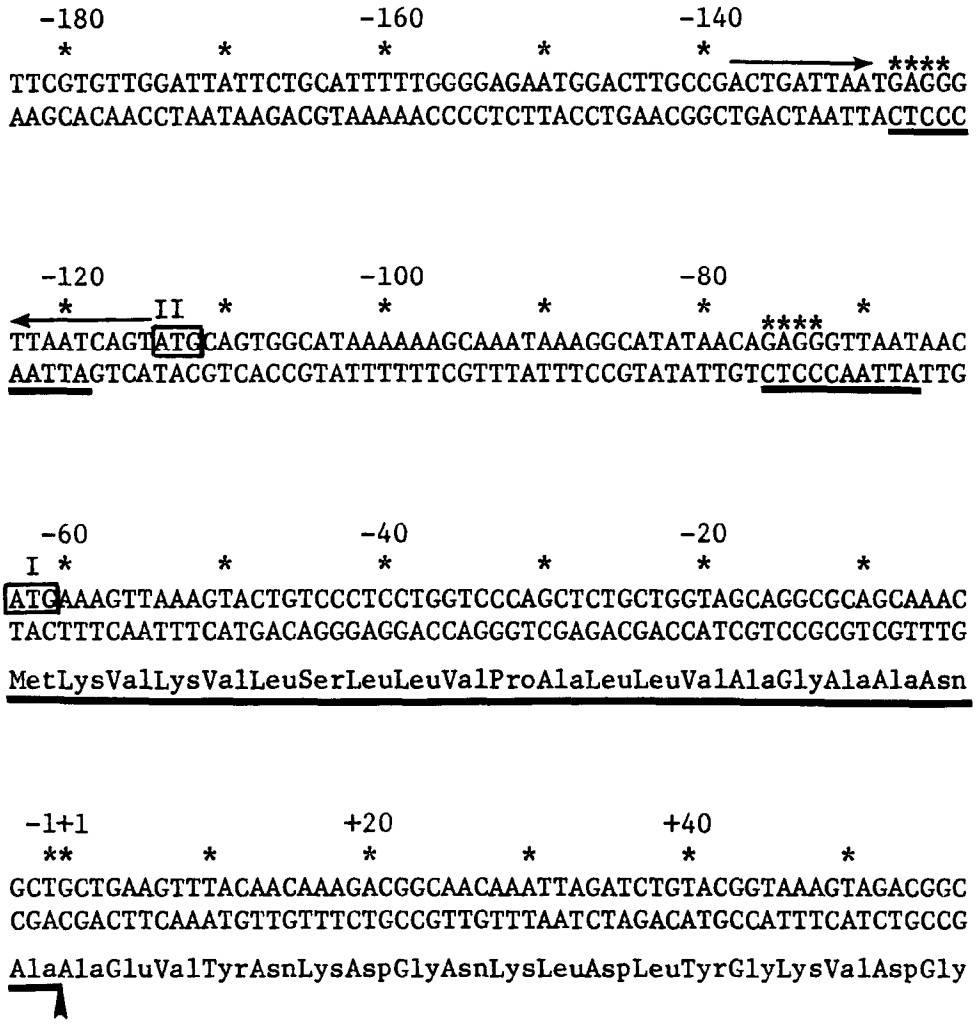

Fig.2. DNA sequence of the 5'-untranslated region and coding region for pro-OmpC protein and the deduced amino acid sequence. The nucleotides are numbered with the first nucleotide of the codon for the NH_2-terminal residue of OmpC protein as +1. The two repetitive sequences including the Shine–Dalgarno sequences (*) are indicated by underlines, both of which are followed by initiation codons (ATG, I or II). Arrows (→ ←) indicate complementary sequences, between which a stem and loop structure ($\Delta G = -8.1$ kcal [24]) can be formed. The signal peptide sequence deduced from the DNA sequence is underlined. The processing site is indicated by a triangle.

ompB strain was used as a recipient for transformation to avoid a possible lethal effect due to an overproduction of OmpC protein, since the *ompC* gene is not expressed in an *ompB* mutant [8]. A clone (pMY111) carrying the 2.7-kb *Hin*dIII fragment was isolated by colony hybridization with the ^{32}P-labeled *Hpa*I–*Eco*RI fragment. When *E. coli* W620 *ompC*::Mu was transformed with the plasmid pMY111, an outer membrane protein with the same M_r-value as OmpC protein was produced in the transformed cells (not shown). Moreover, the cells became sensitive to phage TuIb, which is known to use OmpC protein as its receptor [2]. These results indicate that pMY111 contains the entire *ompC* gene.

3.1. *Identification of the* ompC *gene and its DNA sequence*

To determine the position of the *ompC* gene in the 2.7-kb *Hin*dIII fragment, the restriction map of the fragment was determined as shown in fig.1B. An *Eco*RI site found in the fragment was considered to correspond to the *Eco*RI site of the *Hpa*I–*Eco*RI probe derived from λp10-25. Therefore, we determined the DNA sequence around the *Eco*RI site. We found that sequence on the right hand side of the *Eco*RI site was identical to that of the *Hpa*I–*Eco*RI probe (not shown). To find the coding region for the NH_2-terminal part of OmpC protein, we determined the DNA sequence to the left of the *Eco*RI site (fig.1C). The DNA sequence around the *Bgl*II site was found to code for the amino acid sequence:

Ala–Glu–Val–Tyr–Asn–Lys–Asp–Gly–
–Asn–Lys–Leu–Asp (fig.2)

which is exactly the same as the NH_2-terminal sequence of OmpC protein as determined in [16]. This result confirms that the cloned DNA fragment contains the *ompC* structural gene. Thus, the signal peptide sequence consisting of 21 amino acid residues was also deduced from the DNA sequence as shown in fig.2.

4. DISCUSSION

The NH_2-terminal amino acid sequence of OmpC protein of *E. coli* K-12 has been determined to be:

Ala–Glu–Val–Tyr–Asn–Lys–Asp–Gly–
–Asn–X–Leu–Asp [16]

Fig.3. DNA sequence comparison between the *ompC* gene and the *ompF* gene: (A) DNA sequences encoding the first 15 amino acids of NH_2-terminal residues; (B) the signal peptides of OmpC and OmpF proteins are shown. The *ompF* gene sequence is cited from [6]. Non-homologous regions are connected with empty bars, amino acid differences are underlined.

Exactly the same sequence can be deduced from the DNA sequence (from positions +1 to +36 in fig.2). Since OmpC protein consists of about 340 amino acid residues which corresponds to 1020 base pairs, the 2.7-kb *Hin*dIII fragment should contain the entire *ompC* gene as shown in fig.1.

The DNA sequence also showed that OmpC protein is synthesized as a precursor with an extension at the NH_2-terminal end, consisting of 21 amino acid residues, which appears to have the well-known features of prokaryotic signal peptide sequences (review [18]). These characteristics are as follows:

(i) The existence of basic amino acids near the NH_2-terminus;
(ii) The existence of long hydrophobic residues at the central region;
(iii) The existence of a glycine and a proline residue in the hydrophobic region;
(iv) The existence of the cleavage site on the carboxyl terminal side of an alanine residue.

It is interesting to note that the cleavage site for pro-OmpC protein is Ala–Ala–Glu, which is identical with those for pro-OmpF [6] and pro-PhoE [19] proteins. All of these proteins are known to form diffusion pores across the outer membrane.

The DNA sequence coding for the first 15 amino acids of NH_2-terminal portion of OmpC protein reveals extensive homology with that of the *ompF* gene [6], only 6 out of 45 base pairs are substituted (fig.3). However, there are more drastic changes in the DNA sequences coding for the signal peptides of pro-OmpC and pro-OmpF proteins, including not only single base substitutions but also alterations of more than one base within a codon (fig.3).

More surprisingly, no such homology can be seen in the 5′-untranslated regions of the two proteins. In this region of pro-OmpC protein (fig.2), there are several unique features. The translation initiation codon ($^{-63}$ATG^{-61}) is preceded by a Shine–Dalgarno sequence ($^{-76}$GAGG^{-73}) constituting the ribosome binding site [20]. The distance between these sequences is 9 base pairs. These facts are consistent with the general findings for prokaryotic translation initiation sites (review [21]). It should be pointed out that exactly the same sequence ($^{-76}$GAGGTTAAT^{-67}) which includes the Shine–Dalgarno sequence was found further upstream from positions −128 to −119, and that a stable stem and loop structure can be formed at this region. Moreover, this region is also followed by another $^{-114}$ATG^{-112}. Protein synthesis beginning at this initiation codon would terminate just upstream of the putative ribosome binding site $^{-76}$GAGG^{-73}. These features are summarized in fig.2. The 5′-untranslated region of another major outer membrane protein, the OmpA protein, has similar characteristics [22]. The exact role of such structures is unknown at present. Upstream of this region, there is a possible Pribnow-box (−10 region) $^{-154}$gAgAATG^{-148} (lower case letters denote nucleotides which do not share homology with the consensus sequence) and the RNA polymerase recognition site (−35 region) $^{-179}$TGTTGgattaTT^{-168}, both of which have good homology with the consensus sequences [23].

The expression of the *ompC* and *ompF* genes are under complex regulation exerted mainly by the *ompB* operon. The *ompB* operon has been cloned, and its entire DNA sequence has been determined [4,5]. Since its gene products have been identified [4,5], it is now of great interest to examine how these gene products interact with the promoter regions of the *ompF* and *ompC* genes.

ACKNOWLEDGEMENTS

The authors are grateful to Dr T.J. Silhavy for his kind gift of phage λp10-25. We also which to thank Ms P. Green for her critical reading of this manuscript. This research was supported by grant GM 19043 from the United States Public Health Service and Grant NP3871 from the American Cancer Society. T.M. is on leave from Mitsubishi Kasei Institute of Life Sciences, Japan.

REFERENCES

[1] Inouye, M. (1979) in: Bacterial Outer Membrane; Biogenesis and Function (Inouye, M. ed) pp. 1–12, Wiley, New York.
[2] Reeves, P. (1979) in: Bacterial Outer Membrane; Biogenesis and Function (Inouye, M. ed) pp. 255–291, Wiley, New York.
[3] Kawaji, H., Mizuno, T. and Mizushima, S. (1979) J. Bacteriol. 140, 843–849.
[4] Wurtzel, E.T., Chou, M-Y. and Inouye, M. (1983) J. Biol. Chem. in press.
[5] Mizuno, T., Wurtzel, E.T. and Inouye, M. (1983) J. Biol. Chem. in press.
[6] Mutoh, N., Inokuchi, K. and Mizushima, S. (1982) FEBS Lett. 237, 171–174.

[7] Wurtzel, E.T., Movva, R.N., Ross, F. and Inouye, M. (1981) J. Mol. Appl. Genet. 1, 61–69.
[8] Hall, M.N. and Silhavy, T.J. (1979) J. Bacteriol. 140, 342–350.
[9] Nakamura, K. and Inouye, M. (1979) Cell 18, 1109–1117.
[10] Nakamura, K., Wurtzel, E.T., Pirtle, R.M. and Inouye, M. (1979) J. Bacteriol. 138, 715–720.
[11] Sakano, H., Huppi, K., Heinrich, C. and Tonegawa, S. (1979) Nature 280, 288–294.
[12] Maxam, A.M. and Gilbert, W. (1981) Methods Enzymol. 65, 499–560.
[13] Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560–564.
[14] Southern, E.M. (1975) J. Mol. Biol. 98, 504–517.
[15] Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, R. (1977) J. Mol. Biol. 113, 237–251.
[16] Ichihara, S. and Mizushima, S. (1978) J. Biochem. (Tokyo) 83, 1095–1100.
[17] Inouye, M. and Halegous, S. (1980) CRC Crit. Rev. Biochem. 7, 339–371.
[18] Halegoua, S. and Inouye, M. (1979) in: Bacterial Outer Membrane: Biogenesis and Function (Inouye, M. ed) pp. 67–113, Wiley, New York.
[19] Overbeeke, N., Bergmans, H., Van Mansfeld, F. and Lugtenberg, B. (1983) J. Mol. Biol. in press.
[20] Shine, J. and Dalgarno, U. (1974) Proc. Natl. Acad. Sci. USA 71, 1342–1346.
[21] Gold, L., Pribnow, D., Schneder, T., Shinelding, S., Singer, B.S. and Stormo, G. (1981) Annu. Rev. Microbiol. 35, 365–403.
[22] Movva, N.R., Nakamura, K. and Inouye, M. (1980) J. Mol. Biol. 143, 317–328.
[23] Rosenberg, M. and Court, D. (1979) Annu. Rev. Genet. 13, 319–353.
[24] Salser, W. (1977) Cold Spring Harbor Symp. Quant. Biol. 13, 985–1002.